Virology 422 (2012) 195–204

Contents lists available at SciVerse ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Identification and classification of endogenous retroviruses in the canine genome using degenerative PCR and *in-silico* data analysis

Haiin Jo [a], Hojun Choi [a], Min-Kyeung Choi [a], Ning Song [a], Jin-Hoi Kim [a], Jae-Wook Oh [b], Kunho Seo [c], Han Geuk Seo [a], Taehoon Chun [d], Tae-Hun Kim [e], Chankyu Park [a,*]

[a] Department of Animal Biotechnology, Konkuk University, Hwayang-dong, Kwangjin-gu, Seoul, South Korea
[b] Division of Animal Life Science, Konkuk University, Seoul, South Korea
[c] College of Veterinary Medicine, Konkuk University, Seoul, South Korea
[d] College of Life Sciences and Biotechnology, Korea University, Seoul, South Korea
[e] Animal Genomics and Bioinformatics Division, National Institute of Rural Development Administration Suwon, South Korea

ARTICLE INFO

Article history:
Received 1 July 2011
Returned to author for revision
15 August 2011
Accepted 11 October 2011
Available online 14 November 2011

Keywords:
Endogenous retrovirus
Canine
Genome
Pro
Pol
ERV

ABSTRACT

Pooled genomic DNA from 10 dogs was subjected to polymerase chain reaction with primers targeting the retroviral *pro*/*pol* region. Sequence analysis of 120 clones obtained by PCR revealed 81 of retroviral origin. Subsequent analysis of the dog genome (CanFam 2.0) by BLAST investigation using degenerate PCR products and previously identified retroviral sequences permitted the identification of additional retroviral γ and β sequences. A phylogenetic analysis using the retroviral protease (PR) and reverse transcriptase (RT) sequences in the dog genome resulted in identification of 17 γ and 7 β families. In addition, we also identified 167 spuma-like ERV elements from CanFam 2.0 based on sequence homology to murine (Mu)ERV-L and human (H)ERV-L. Our results could contribute to the understanding of the influence of retroviruses in shaping the genome structure and altering gene expression by providing quantitative and locational information of ERV loci and their diversity in the dog genome.

Introduction

Retroviruses can be broadly classified into two groups: simple (alpha, beta, gamma and epsilon retroviruses) and complex (lenti-, delta- and spuma-viruses). Only simple retroviruses are capable of being endogenous in their host genome with the exception of spuma-viruses (Weiss, 2006). Endogenous retroviruses (ERVs) integrate into the host germ line in a Mendelian order (Gifford and Tristem, 2003). Their typical structure consists of *gag* (viral core proteins), *pro–pol* (viral enzymes), and *env* (envelope proteins) genes with long terminal repeats (LTRs) at their 5′ and 3′ termini (Patience et al., 2001).

It has been proposed that ERVs co-evolved with their host genome (Britten, 1996; McDonald, 1993) and similar retroviral sequences in diverse species suggest cross-species transmission events (Benit et al., 2001). As a result, ERVs carry a significant disposition of numerous point mutations, deletions, and insertions over time, and thus act as a 'fossil record' for understanding retroviruses and the co-evolution of their host genome (Coffin et al., 1997). Therefore, ERV-originated sequences are sometimes difficult to recognize due to severe deterioration as a result of accumulated mutations.

In most cases, ERVs no longer maintain their activity, but ERV insertions produce diverse effects on their host genome including altered gene expression by causing ectopic recombination (Hughes and Coffin, 2005), premature polyadenylation (Palmarini et al., 2002), or aberrant splicing (Maksakova et al., 2006). Therefore, the changes in the genome caused by ERVs can result in diseases in organisms (Maksakova et al., 2006; Ryan, 2004).

ERVs have been studied in diverse species including mammals (Tristem, 1996; Tristem et al., 1996), chickens (Dunwiddie et al., 1986), reptiles, amphibians, and fish (Herniou et al., 1998). In mammals, detailed genome level characterizations of ERVs were carried out in several species including humans (Lower et al., 1996), chimpanzees (Polavarapu et al., 2006), swine (Klymiuk et al., 2002; Patience et al., 2001), sheep (DeMartini et al., 2003; Klymiuk et al., 2003), and cattle (Garcia-Etxebarria and Jugo, 2010; Xiao et al., 2008b). From these studies, three classes of ERVS were reported: Class I (γ [C-type]), Class II (β [B/D-type]), and Class III (spuma-like).

* Corresponding author at: Department of Animal Biotechnology, Konkuk University, Hwayang-dong, Kwangjin-gu, Seoul 143–701, South Korea. Fax: +82 2 457 8488.
E-mail addresses: hi458@konkuk.ac.kr (H. Jo), shukle2@konkuk.ac.kr (H. Choi), sirius1029@nate.com (M.-K. Choi), funny-sonia@163.com (N. Song), jhkim541@konkuk.ac.kr (J.-H. Kim), ohjw@konkuk.ac.kr (J.-W. Oh), bracstu3@konkuk.ac.kr (K. Seo), tchun@korea.ac.kr (T. Chun), thkim63@korea.kr (T.-H. Kim), chankyu@konkuk.ac.kr (C. Park).

0042-6822/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2011.10.010

The dog is an important animal species as a major companion animal (Clutton-Brock, 1995; Schwartz, 1997) and disease model (Patterson et al., 1982; Patterson et al., 1988; Wayne and Ostrander, 1999). The current dog genome assembly, CanFam 2.0, covers 98% of the entire canine genome with 2400 Mb of assembled sequences (Lindblad-Toh et al., 2005), and the brief estimation of ERV sequences in the dog genome have been attempted (Blikstad et al., 2008; Lindblad-Toh et al., 2005). Very recently, Martínez Barrio et al. (2011) reported a comprehensive analysis on the complexity and integration pattern of canine endogenous retroviruses (CfERVs) using an ERV analysis program. The discovery and detailed characterization of ERVs in the canine genome are important to improve annotation of repetitive elements and facilitate an understanding of the role of ERVs in modification of the genome (Jern et al., 2005; Lower et al., 1996).

A limited number of ERV sequences were initially identified from experimental approaches including the analysis of cDNA-library clones (La Mantia et al., 1991). More recently, a degenerate polymerase chain reaction (PCR) method using specific primers for the conserved active site motifs of protease (PR) and reverse transcriptase (RT) has been successfully employed to study diverse ERVs in the genome (Akiyoshi et al., 1998; Klymiuk et al., 2002; Xiao et al., 2008b).

In this study, we first amplified the conserved *pro*/*pol* region of canine ERVs using degenerate PCRs and analyzed the products after cloning and sequencing. Subsequent *in-silico* searches of ERV sequences against the reference dog genome using the information derived from the degenerate PCRs and previously reported retroviral sequences resulted in a genome level evaluation of ERV elements which we classified into individual ERV families on the basis of *pro*/*pol* sequence diversity.

Results

Identification of ERV gamma (γ) elements using degenerate PCR

Degenerative PCR using *pro*/*pol* specific primers and dog genomic DNA and subsequent cloning generated a total of 120 candidates with possible PR and/or RT containing sequences. Sequence analysis of the clones revealed 86 (71.7%) with a length of 0.6 to 1.0 kb having significant sequence similarity (40–97%) to previously identified γ retroviruses (data not shown). Among clones of retroviral origin, 10 clones of five different pairs had repetitious sequences, showing a result similar to previous studies (Patience et al., 2001; Xiao et al., 2008b). Therefore, we identified 81 unique sequences with retroviral origins using degenerate PCR.

The clones (or elements in *in-silico* analysis) having more than 80% nucleotide sequence identity were clustered as a family into four groups with 48, 25, 6, and 2 clones each. The most abundant group was designated as *Canis familiaris* endogenous retrovirus (CfERV) γ1, followed by γ2, γ3, and γ4. Their phylogenetic relationship showed the same topology as the order of clone abundance, thus the genetic distance between CfERV γ1 and γ4 was furthest (Supplementary Fig. 1A). The sequence having the fewest nonsense mutations and being most closely related to potentially infectious ERVs was determined as the representative for each family. Phylogenetic analysis was performed together with previously identified γ- and β- retroviruses. All sequences from degenerate PCR were clustered with gammaretroviruses and no clones clustered with betaretroviruses (Supplementary Fig. 1B).

The CfERV γ2 sequences were relatively shorter (557–594 base pair [bp]) than others (γ1, γ3, and γ4) which ranged from 605 to 946 bp. Subsequent analysis showed that CfERV γ2 contained the PR region but lacked most RT sequences. However, CfERV γ2 clones should contain at least a part of the RT specific motif sequence to be amplified by *pro*/*pol* specific primers, which was supported by confirming the presence of both PR and RT sequences in the primer site and RNaseH sequences from the sequencing results of γ2 clones (data not shown).

In-silico analysis of dog ERVs from CanFam 2.0 detecting 176 ERV γ and 8 β elements

Basic Local Alignment Search Tool (BLAST) analysis using 81 *pro*/*pol* sequences from degenerative PCR through a NCBI blastn suite, BLAST Dog Sequences, identified 160 loci with ERV γ elements (Supplementary Table 1). To identify β family elements, we performed BLAST analysis in the same way using nine previously identified betaretroviruses (TvERV, SRV, JSRV, OERV β1, OERV β2, OERV β3, HERV-K, MPMV, PERV β5) of several species. Although we used diverse β retroviral sequences, we only found 8 BLAST matches from the entire canine genome (Supplementary Table 2), indicating a limited presence of ERV β elements in the dog genome. Since ERVs are a member of long terminal repeat (LTR) retrotransposons, we tried to identify full-length LTR retrotransposons with both 5′ and 3′ LTRs. Although 557 putative full-length LTR retrotransposons were identified from the LTR_STRUC (McCarthy and McDonald, 2003) analysis, only 39 were confirmed as *pro*/*pol* containing elements (data not shown). Most sequences were retrotransposons with LTRs at both the 5′ and 3′ ends without any significant matches to retroviral sequences. We also attempted to identify ERV sequences using three reported ERV sequences of 7000 bp of *Canis familiaris* origin from the Repbase (http://www.girinst.org.repbase/). We combined all ERV sequences from different approaches, removed redundant or severely mutated sequences including large deletions or insertions which might cause problems in constructing a phylogenetic tree, and chose sequences which contained the *pro*/*pol* region and mapped to different locations in CanFam 2.0. Finally, 173 ERV γ and 8 β elements were identified from the current dog genome assembly (Supplementary Tables 3 and 4). Each element was analyzed for chromosomal coordinates within the genome assembly, element sizes (bp) and length, and sequence identity between 5′ and 3′ LTRs. The size of ERV elements was estimated by counting the number of nucleotides between the start of the 5′ LTR and the last nucleotide of the 3′ LTR. Thus, element length was defined only for those (71.8%) with recognizable LTR sequences at both ends.

In addition, we evaluated 407 CfERV elements reported by Martínez Barrio et al. (2011) according to our ERV determination criteria which were described in Materials and methods. A significant number of CfERV elements defined by them did not meet our criteria and were excluded from our result. Consequently, three γ elements were additionally added to PR/RT containing CfERVs, resulting in the identification of 176 γ and 8 β ERV elements from the dog genome (Supplementary Tables 3 and 4).

A total of 74.4% (n = 131) of ERV γ and 25% (n = 2) of β elements had two recognizable LTRs, while 4.5% (n = 8) of γ elements were solo LTR ERVs; therefore, we were not able to identify LTR sequences from 21.0% (n = 37) and 75% (n = 6) of γ and β elements, respectively. The average size of ERV elements was 8427 bp (range 8259 to 8594) for CfERV β and 6380 bp (3555 to 18462) for γ elements. These results are consistent with those of previous studies from other species (Bowen and McDonald, 1999; Jern et al., 2005; McCarthy and McDonald, 2004; McCarthy et al., 2002). The large variation in size among γ elements was due to the presence of nucleotide deletions and insertions within the elements. For example, the element *Der10-5* (CfERV γ10, 3979 bp) had Gag and PR but lacked RT between 5′ and 3′ LTRs (Supplementary Table 6). In contrast, we found insertions of both transposase and non-annotatable nucleotide sequences within the element *Der18-7* (CfERV γ10, 14434 bp).

To estimate the total amount of ERV sequences in the dog genome, we simply multiplied the copy numbers of β and γ elements by the average size of element length for each genus because the exact boundary of ERV elements with missing or solo LTRs was uncertain.

In addition, we also counted nucleotide numbers of actual matches to previously known ERVs using 139 ERV-related elements without PR or RT, which summed to approximately 2.02 Mb or 0.084% of the canine genome (Supplementary Table 5). The estimated value for all ERV elements in the canine genome from our analysis was equal to 3.02 Mb or 0.125% of the dog genome (Supplementary Table 5).

Classification of gammaretroviruses in the dog genome into 17 subgroups (CfERVγ1 to γ17)

To represent all canine ERV γ elements, those identified from degenerate PCR and *in-silico* analysis were combined and used for family classification. Based on the structural integrity of the *pro*/*pol* region, which typically contains aspartyl PR, RT, RnaseH, and integrase (Vogt, 1997), the ERV elements were divided into three groups, depending on the presence of both PR and RT, PR only, or RT only, and identified as the PR–RT, PR, and RT groups, respectively. Among ERV γ elements, the PR group was most frequent (n = 84), followed by PR–RT (n = 70) and RT (n = 19).

The ideal way to classify the family of ERV elements is to perform phylogenetic analysis using all available sequences together. However, because phylogenetic analysis using sequences consisting of partially different regions could be problematic, we classified each group separately and treated them as different families. As a result, 173 CfERV γ elements were classified into 17 CfERV γ families (Figs. 1 and 2).

In degenerate PCR, CfERV γ1 was the largest family group followed by CfERV γ2 (Supplementary Fig. 1A). However, the combined analysis of ERV elements from both degenerate PCR and *in-silico* analysis showed that CfERV γ10 which lacks RT sequences possessed the largest number (n = 77) of elements and CfERV γ1 which is the PR–RT group was second (n = 54). Although CfERV γ10 was the most abundant ERVs in the dog genome (Fig. 2A), they were undetectable by degenerate PCR due to the lack of the RT sequence or 3′ primer binding site, which was confirmed by the *in silico* analysis of CfERV γ10 elements in CanFam2.0. In addition, phylogenetic analysis of the *in-silico* derived PR–RT containing elements (n = 70) were classified into CfERV γ1 and γ3 to γ8 families with 54, 7, 2, 1, 1, 2, and 3 elements, respectively (Supplementary Fig. 2).

CfERV γ2 consisted of 25 sequences from only degenerative PCR. There were no ERV elements belonging to the CfERVγ2 family from CanFam 2.0, indicating the presence of individual or breed dependent CfERV diversity because ERV elements from degenerative PCR resulted from the pooled DNA of 10 dogs, while those from CanFam 2.0 were from an individual genome of a Boxer breed dog. However, although CfERV γ2 were classified into different families due to the presence of partial RT sequences, they showed the closest relationship in their sequence and structure with CfERV γ9 and γ10 suggesting that variation at the family level was very limited.

CfERV γ9 and γ10 consisted of elements belonging to the PR group with 7 and 77 elements each (Fig. 2A). A total of 19 ERV γ elements belonging to the RT group were classified into CfERV γ11 to γ17 with 6, 1, 3, 1, 1, 1, and 6 elements, respectively (Fig. 2B). Interestingly, all CfERV γ11 elements were present in chromosome 26 and their chromosome coordinates were close to each other, suggesting either recent duplication or integration events.

The average LTR length of CfERV γ1 elements was 416 bp (77 to 531) which was similar to that of other species including chickens (200 to 350 bp; Huda et al., 2008), mice (300 to 600; McCarthy and McDonald, 2004), drosophila (260 to 690; Bowen and McDonald, 2001), chimpanzees (300 to 900; Polavarapu et al., 2006) and cattle BERV γ4 (469 to 470; Xiao et al., 2008a). The LTR length of other CfERV γ families was similar to that of CfERV γ1 (Supplementary Table 3).

Classification of betaretrovirus in the dog genome into subgroups (CfERV β1 to β7)

Phylogenetic analysis was carried out to determine the family relationship among eight ERV β elements which were classified into β1 to β7 families based on genetic distance and structural differences (Fig. 3). CfERV β1 to β5 were the PR–RT group and β6 and β7 were the RT group; therefore, CfERV β6 and β7 were not included in the phylogenetic tree constructed with β elements of the PR–RT group. Only CfERV β1 consisted of more than one sequence (*Derb-31* and *Derb-32*) and others were single-element families. Among eight CfERV β elements, two had both 5′ and 3′ LTRs and thus were used to determine the average size of CfERV β elements (8427 bp) and LTR lengths (377 bp).

Conservation of the pro/pol specific motif sequences among CfERV elements

A total of 181 CfERV γ and β elements from *in-silico* analysis were evaluated for conservation or sequence variations of PR–RT specific motifs (Supplementary Tables 6 and 7). CfERVγ1, a member of the

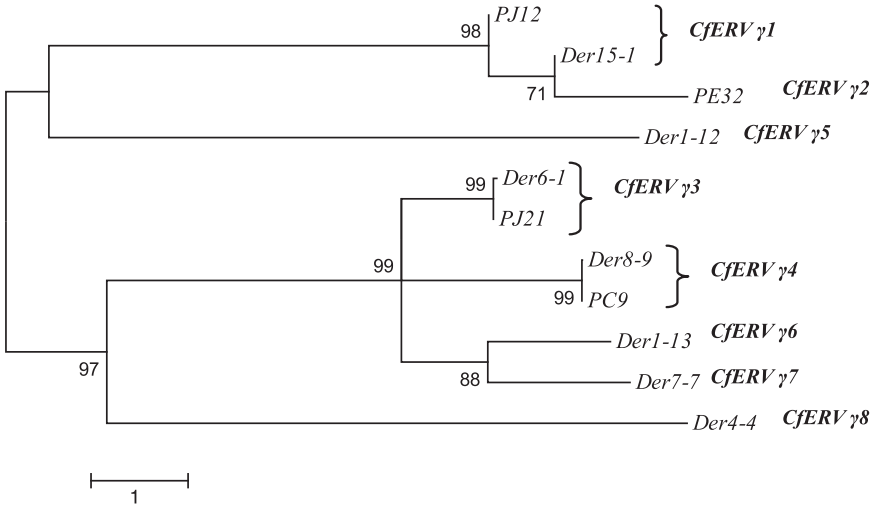


Fig. 1. Phylogenetic analysis and family classification of PR and RT-containing CfERV γ elements. A neighbor-joining tree was constructed using 11 representative PR–RT-group ERVγ elements including degenerative PCR (n = 4) and *in-silico* analysis (n = 7) of CanFam 2.0. Three *in-silico* identified elements (Der15-1, Der6-1, Der8-9) were clustered together with elements from degenerate PCR (PJ12, PE32, PJ21, and PC9). Family names are shown in bold letters. The four new family names (CfERV γ5 to γ8) were assigned from the *in-silico* identified CfERV γ elements. Numbers at the branch nodes denote bootstrap values (>50) from 1000 replicates.

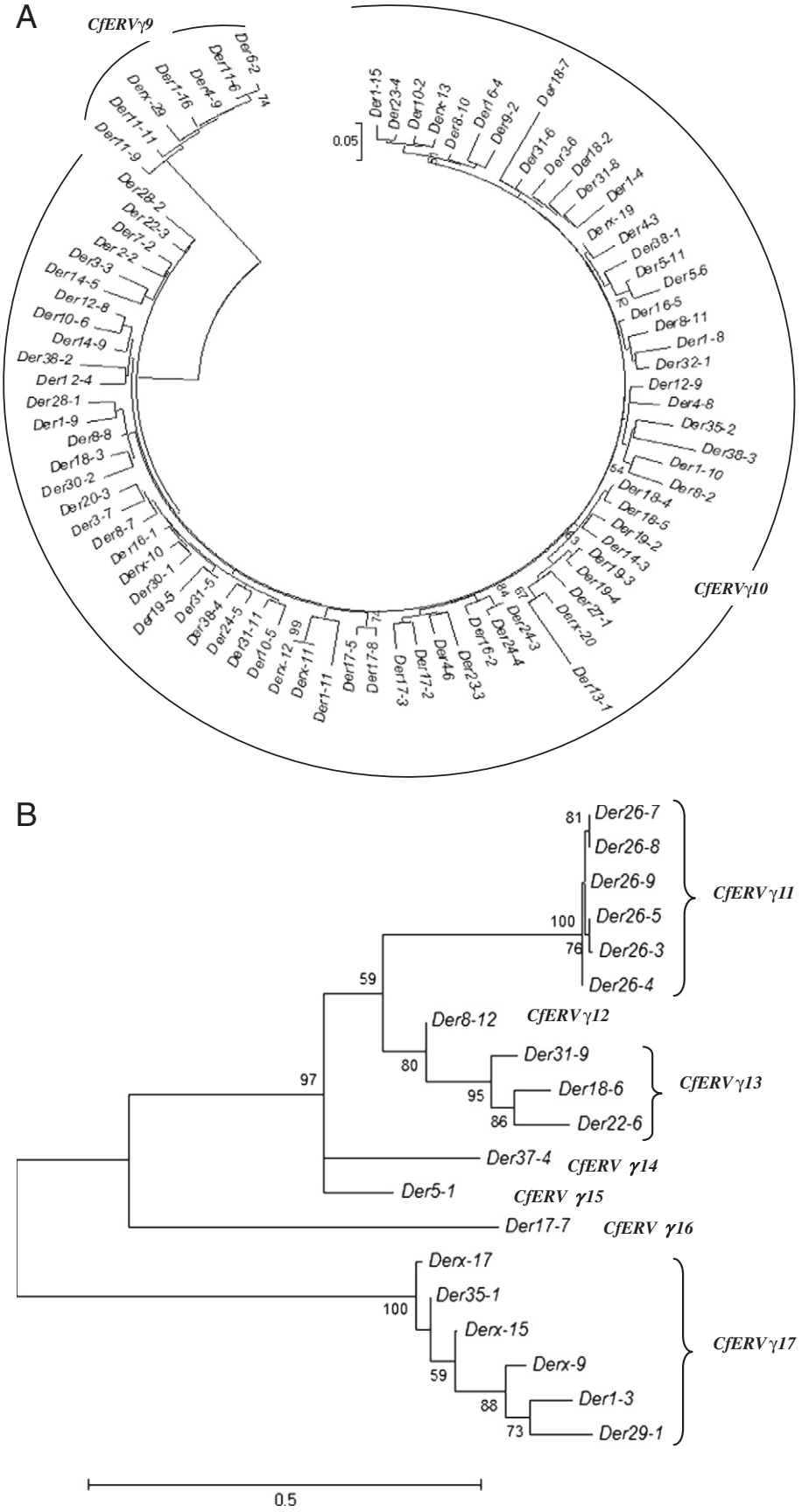


Fig. 2. Phylogenetic analysis and family classification of *in-silico* identified CfERV γ elements belonging to the PR and the RT groups. Neighbor-joining trees were constructed with 84 PR-group CfERV elements (A) and 19 RT-group CfERV elements (B). Family names for the PR-group elements were assigned as CfERV γ9 and γ10 and for RT group as CfERV γ11 to γ17. Numbers at the branch nodes denote the bootstrap values (>50) from 1000 replicates.

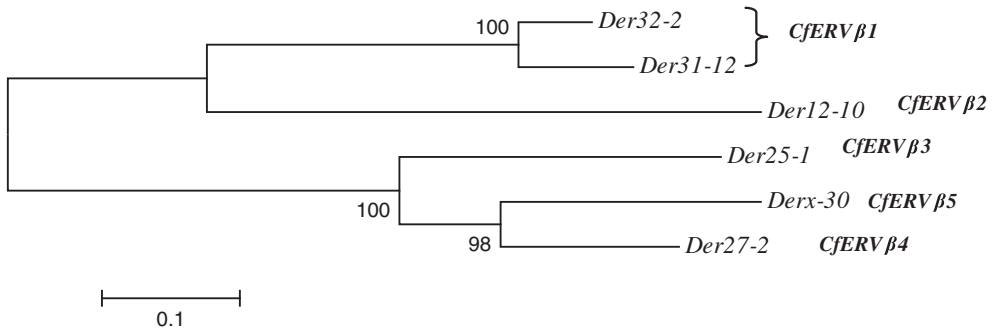


Fig. 3. Phylogenetic analysis and family classification using *in-silico* identified PR–RT-containing CfERV β elements. A neighbor-joining tree was constructed with 6 CfERV β elements. Family names were assigned as CfERV β1 and β5. Numbers at the branch nodes denote the bootstrap values (>50) from 1000 replicates.

PR–RT group, showed intact motif conservation for both PR (DTGA) and RT (LPQGFKN and YVDD) regions. All PR group elements (CfERV γ9 and γ10) also showed conservation of the PR specific motif sequence, DTGA, with only slight variations. In the RT group, except for CfERV γ17 elements which showed well conserved RT motif sequences, LPQGFKN and YVDD, other groups including CfERV γ11 to γ16 showed relatively severe variations in the sequences. Grouping of gamma- and betaretroviruses was performed according to differences in RT-specific motif sequences between them. CfERV γ elements had LPQGF(R/K)(D/N) and YVDD motifs, while β elements had L(P/A)QGM(A/V)N and highly variable YMDD motifs (Fig. 4B and Supplementary Table 7), which may be the reason for amplification failure of the CfERV β elements from degenerate PCR.

Identification of 167 spuma-like ERV elements from CanFam 2.0

The presence of spuma-like ERV (CfERV-L) in the canine genome was briefly described in supplementary data of the canine genome project (Lindblad-Toh et al., 2005). For more detailed analysis, NCBI BLAST analysis using two spuma-like ERV mouse and human elements, MuERV-L and HERV-L, was performed and ~738 matches in the dog genome were identified (data not shown). Among those matches, 167 elements were categorized into PR–RT (n=100), PR (n=25), and RT (n=42) groups (Supplementary Table 8).

Old ERVs such as HERV-L and MALR elements have been described as difficult to classify into subgroups (Blomberg et al., 2009). Consistent with this information, most of the elements were not clustered and separated into single element clusters when we attempted to assign family classifications to CfERV-L elements using phylogenetic analysis (data not shown). Only two pairs of CfERV-L elements among 167 showed >90% sequence identity between them. Therefore, we were not able to make successful family classifications with CfERV-L elements although the number was similar to that of γ elements. The averaged pairwise sequence similarity values within each genus were 0.69, 0.49, and 0.43 for CfERV γ, -L, and β elements, respectively.

We also attempted to identify LTR sequences from CfERV-L elements using the same method which we used for identifying the LTRs from CfERV β and γ elements. However, we were not able to determine regions with clear characteristics of retroviral LTRs. Therefore, to estimate the total amount of CfERV-L elements in CanFam 2.0, we simply summed the size of matched nucleotide sequences to MuERV-L and HERV-L in CanFam 2.0. A total of 738 CfERV-L elements occupied approximately 0.042% or 1.0 Mb of the dog genome (Supplementary Table 5).

Chromosomal distribution of ERV γ and β elements in the dog genome

Fig. 5 shows the distribution of ERVs in CanFam 2.0 for 38 dog autosomes and chromosome X. A total of 175 ERV γ, 8 β, and 167 spuma-like (CfERV-L) elements distributed unevenly through different chromosomes although larger chromosomes tend to have more ERVs. The highest insertion frequency of ERVs was observed from chromosome X with 20 γ and 1 β element. The autosome having the highest insertion frequency was chromosome 1 with 15 γ elements followed by chromosome 8 with 12 γ elements. We were not able to find any PR- or RT-containing ERV γ and β elements from chromosomes 21 and 33. However, considering CfERV-L as well as *gag* or *env* containing loci without PR or RT, every chromosome in the dog genome contained at least a few ERV related sequences (Supplementary Tables 8).

Estimation of ERV integration time in the dog genome

The age of a provirus can be estimated from sequence comparison between flanking 5′ and 3′ LTRs of ERV sequences because both LTRs are identical in their sequences at the time of provirus formation and, without selective pressure, both LTRs independently accumulate mutations over time (Dangel et al., 1995; Johnson and Coffin, 1999; Lebedev et al., 2000). The nucleotide substitution rate per year for pseudogenes and noncoding regions in primates was calculated to be $2.3–5.0\times10^{-9}$ (Johnson and Coffin, 1999) and has been used to estimate the retroviral integration time in mammals (Garcia-Etxebarria and Jugo, 2010; Tonjes and Niebert, 2003; Xiao et al., 2008a). To minimize outlier effects, sequence identities <75% and length difference >5 bp between 5′ and 3′ LTRs were excluded. We estimated that the age of CfERV γ1 (n=16) to range from 25.63 to 4.8 million years (MY) and γ10 (n=27) from 24.85 to 2.35 MY (Table 1).

We identified a total of 132 CfERV elements with both 5′ and 3′ LTRs from the analysis of γ and β elements in CanFam 2.0. To evaluate the age distribution of all available CfERV elements in the dog genome, we divided them into four groups depending on their 5′ and 3′ LTR sequence identity as follows: <75, 75 to <85, 85 to 95, and >95%. The group with 85 to 95% sequence identity was the largest (50.4%, n=66), followed by >95% (29.0%, n=38), 75 to <85% (16.8%, n=22), and <75% (3.8%, n=5) when we used all available ERV elements. The proviral integration time for CfERV γ elements ranged between approximately 0.83 and 42.97 MY (Table 2).

Discussion

We attempted to identify ERV elements in the dog genome, carried out subgroup classifications, and identified their characteristics and genome distribution. We particularly focused on the ERV elements which contained *pro*/*pol* sequences regardless of other structural retrovirus elements. While our analysis procedures relying on manual evaluation of genome BLAST analysis were labor intensive, this enabled us to focus only on sequences with clear evidence of ERV elements.

Several previous studies classified ERVs on the basis of full-length characteristics (Huda et al., 2008; McCarthy et al., 2002; Polavarapu et al., 2006). In this study, we used the retroviral PR and RT regions

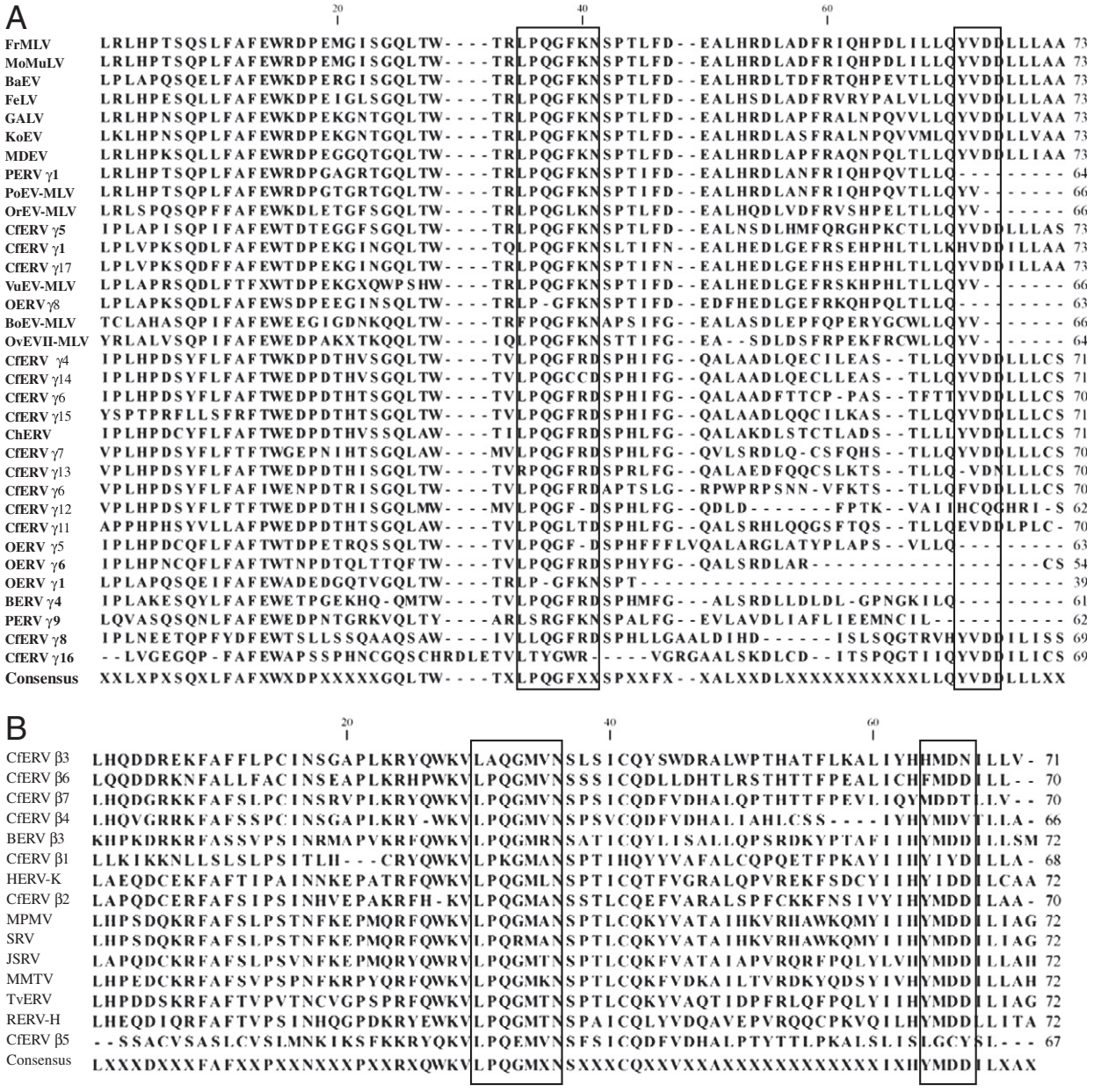
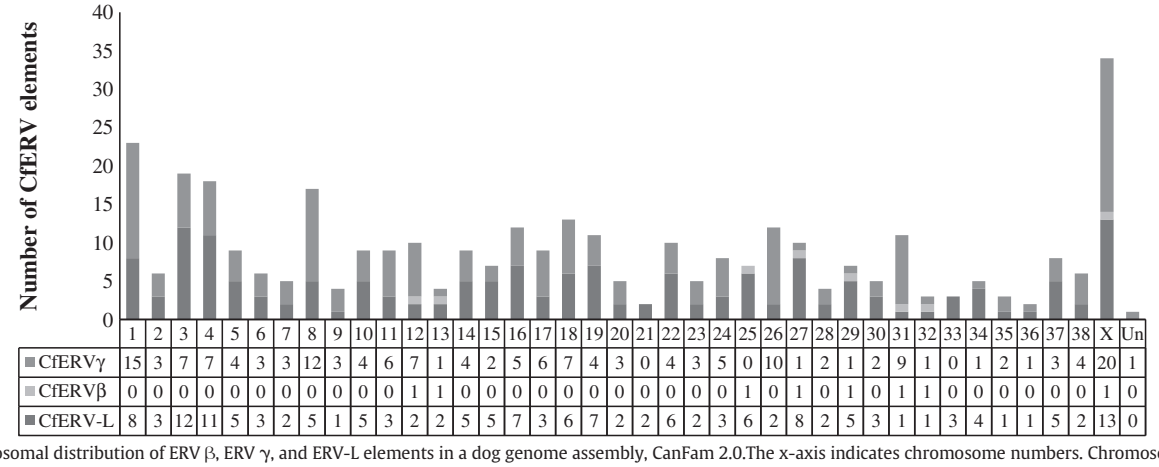


Fig. 4. Comparison of the amino acid sequences of the RT region among CfERV γ and β elements. RT region sequences with ~73 amino acids from gamma- (A) and betaretroviruses (B) including CfERV γ and β elements were aligned using CLUSTALW2. A dash (−) denotes absence of an amino acid. Consensus amino acids were assigned on the basis of 60% amino acid frequency for a given site. "X" in the consensus sequences indicates absence of consensus. The boxed regions indicate specific RT motif sequences.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	X	Un
CfERVγ	15	3	7	7	4	3	3	12	3	4	6	7	1	4	2	5	6	7	4	3	0	4	3	5	0	10	1	2	1	2	9	1	0	1	2	1	3	4	20	1
CfERVβ	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	0	0	0	0	0	0	1	0	
CfERV-L	8	3	12	11	5	3	2	5	1	5	3	2	2	5	5	7	3	6	7	2	2	6	2	3	6	2	8	2	5	3	1	1	3	4	1	1	5	2	13	0

Fig. 5. Chromosomal distribution of ERV β, ERV γ, and ERV-L elements in a dog genome assembly, CanFam 2.0.The x-axis indicates chromosome numbers. Chromosome X (n = 34, 9.8%) harbors the largest number of CfERV elements followed by chromosome 1 (n = 23, 6.6%). Chromosomes 21 and 33 have no CfERV β and γ elements.

Table 1
Estimation of the retrovirus integration time to the dog genome based on long terminal repeat (LTR) sequence divergence.

CfERV family	No. of ERVs	Nucleotide divergence (D)	Substitution rate (R) $\times 10^{-9}$	Approx. age (T) ($\times 10^6$ years old)	
				Min	Max
CfERV γ1	16	0.083 ± 0.035	2.3	10.43	25.63
			5	4.80	11.79
CfERV γ10	27	0.069 ± 0.045	2.3	5.11	24.85
			5	2.35	11.43

Note: CfERV γ1 and γ10 elements were the most abundant CfERV families among protease-reverse transcriptase (PR–RT) and PR group elements, respectively. T = D/2R. "D" indicates the nucleotide divergence between the two LTR sequences which were estimated by the Kimura-2-parameter method using MEGA version 4. "R" indicates the estimated average substitution rate per nucleotide per year (Johnson and Coffin, 1999; Kimura, 1980). Endogenous retroviral (ERV) elements with sequence identity <75% and >5 base pair (bp) length difference between 5′ and 3′ LTRs were excluded from the analysis to decrease the influence of outliers.

for subgroup classification of retroviral sequences because PR and RT are the most conserved regions but still contain enough sequence diversity among different ERV families (Gifford and Tristem, 2003). However, since our initial analysis was based on the presence of the PR and RT sequences for identifying and characterizing ERV elements in the dog genome, this might unintentionally lead to elimination of ERV elements without complete PR and RT regions in our analysis. For example, sequences containing only a partial PR–RT region, such as RnaseH or integrase, could be excluded in our analysis. Consistent to this, our additional searches for *gag*- or *env*-containing sequences devoid of the PR or RT motif identified 139 elements in CanFam 2.0 which were then not suitable for the family classification of CfERVs (Supplementary Table 5).

To compare the structural differences of CfERV elements in the dog genome, we analyzed their substructural elements using BLASTX (Supplementary Fig. 3). The *gag-pro-pol* structure was the most frequent (58.1%, n = 107) among 184 CfERVs, followed by the more complete *gag-pro-pol-env* structure (17.7%, n = 32). A total of 15 CfERV γ and 2 CfERV β elements maintained the complete retroviral structure (9.5%, n = 17) as the *LTR–gag–pro–pol–env–LTR* (Supplementary Tables 3 and 4), while others included *gag–pol, gag–pro, pol–env, pro* only, *and pro–pol–env* structures in the order of their abundance.

When we compared the abundance of genetic elements involved in the proviral structure of ERVs, the number of *pol* containing CfERVs were highest at 172, followed by *gag* and *pro* with 163 and 162, respectively. However, the number of *env*-containing elements was only 38, which was much smaller than other structural gene containing groups.

Table 2
Distribution of the retrovirus integration time in the dog genome based on LTR sequence divergence.

Sequence identity between 5′/3′ LTRs (%)	No. of ERVs	Nucleotide divergence (D)	Substitution rate (R) $\times 10^{-9}$	Approx. age (T) ($\times 10^6$ years old)	
				Min	Max
75 ≤ x < 85	3	0.181 ± 0.017	2.3	35.58	42.97
			5	16.37	19.77
85 ≤ x < 95	40	0.084 ± 0.026	2.3	12.49	23.99
			5	5.74	11.04
95 ≤ x < 100	27	0.024 ± 0.016	2.3	1.81	8.81
			5	0.83	4.05

Note: T = D/2R. "D" indicates the nucleotide divergence between the two LTR sequences which were estimated by Kimura-2-parameter method using MEGA version 4. "R" indicates the estimated average substitution rate per nucleotide per year (Johnson and Coffin, 1999). ERV elements with sequence identity <75% and >5 bp of length difference between 5′ and 3′ LTRs were excluded from the analysis to decrease the influence of outliers.

The dog family, Canidae, contains 34 closely related species including the red fox which diverged within the last ~10 million years (Lindblad-Toh et al., 2005). A close phylogenetic relationship between CfERV γ1 and the Vulpus (red fox) ERV, VuEV-MLV, with 85% sequence identity exists, suggesting that CfERV γ1 and VuEV-MLV might stem from the same origin (Fig. 6). This is consistent with the proviral integration time of CFERV γ1 from 25.63 to 4.8 MY which could precede the divergence time of dogs and red foxes from a common ancestor.

It is quite challenging to accurately define degenerated proviral ERV sequences in mammalian gnomes. Estimations of the total amount of ERV sequences in the genome vary significantly even for the same species depending on the definition of ERV elements or analysis methods used (Blikstad et al., 2008; Lindblad-Toh et al., 2005). The estimated value (0.125%) for all ERV elements in the canine genome from our study was slightly lower than 0.15% from the RetroTector analysis (Martínez Barrio et al., 2011). This discrepancy may indicate the underestimation of ERVs in the canine genome in our study or be due to our strict analysis criteria in evaluating ERV elements. A number of sequences classified as CfERV elements from RetroTector analysis by Martínez Barrio et al. (2011) were classified to either LINE-1, non-LTR retrotransposon or non-ERV related sequences in our view and therefore excluded from our results. As an additional difference, our analysis identified significantly more spuma-like ERV elements (CfERV-L) in the dog genome (Supplementary Tables 5 and 8) than the study of Martínez Barrio et al. (2011) in which they reported only a few spuma-like ERV elements. The result of the dog genome consortium (Lindblad-Toh et al., 2005) also showed that the dog genome contains more ERV-L than other types of ERV elements.

The analysis of data from multiple individuals is likely to be more comprehensive than the use of a single individual in representing diverse types of ERV elements for the species. Therefore, the analysis of ERV elements using degenerate PCR with mixed DNA from 10 dogs in this study could also help to evaluate the individual diversity of CfERVs although we did not address it in this study.

Conclusion

We performed the genome-level identification and characterization of canine ERV elements. Our results are consistent with a previous analysis which suggested that the amount of retroviral elements in the canine genome is relatively smaller than several other species including human, mouse, and zebrafish (Blikstad et al., 2008). We were not able to identify any CfERV elements with intact open reading frames for all functional retroviral elements in CanFam 2.0. Although we did not attempt to analyze the functional importance of these elements in the canine genome, this study provides a starting point to begin to answer such questions.

Materials and methods

Cloning of retroviral pro/pol sequences from the dog genome by degenerate PCR

Genomic DNA was prepared from ~0.5 g tissue samples by a simple lysis method (Miller et al., 1988). An equal amount of DNA from 10 crossbred dogs was pooled as template DNA for PCR. PCR parameters consisted of 2 min at 80 °C followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 43 °C for 30 s, extension at 74 °C for 1 min, and finally one cycle at 94 °C for 30 s, 43 °C for 3 min, and 74 °C for 10 min. Amplifications were performed in 25 μl reactions containing 40 pmol of each primer, 200 μM dNTPs, PCR buffer (10 mM Tris [pH8.3], 50 mM KCl, 1.5 mM $MgCl_2$), 100 ng of genomic DNA and 2 U of Taq polymerase (Supertherm; Roche, Mannheim, Germany). We utilized six pairs of degenerate primers (Herniou et al., 1998; Tristem et al., 1996) consisted of two forward primers,

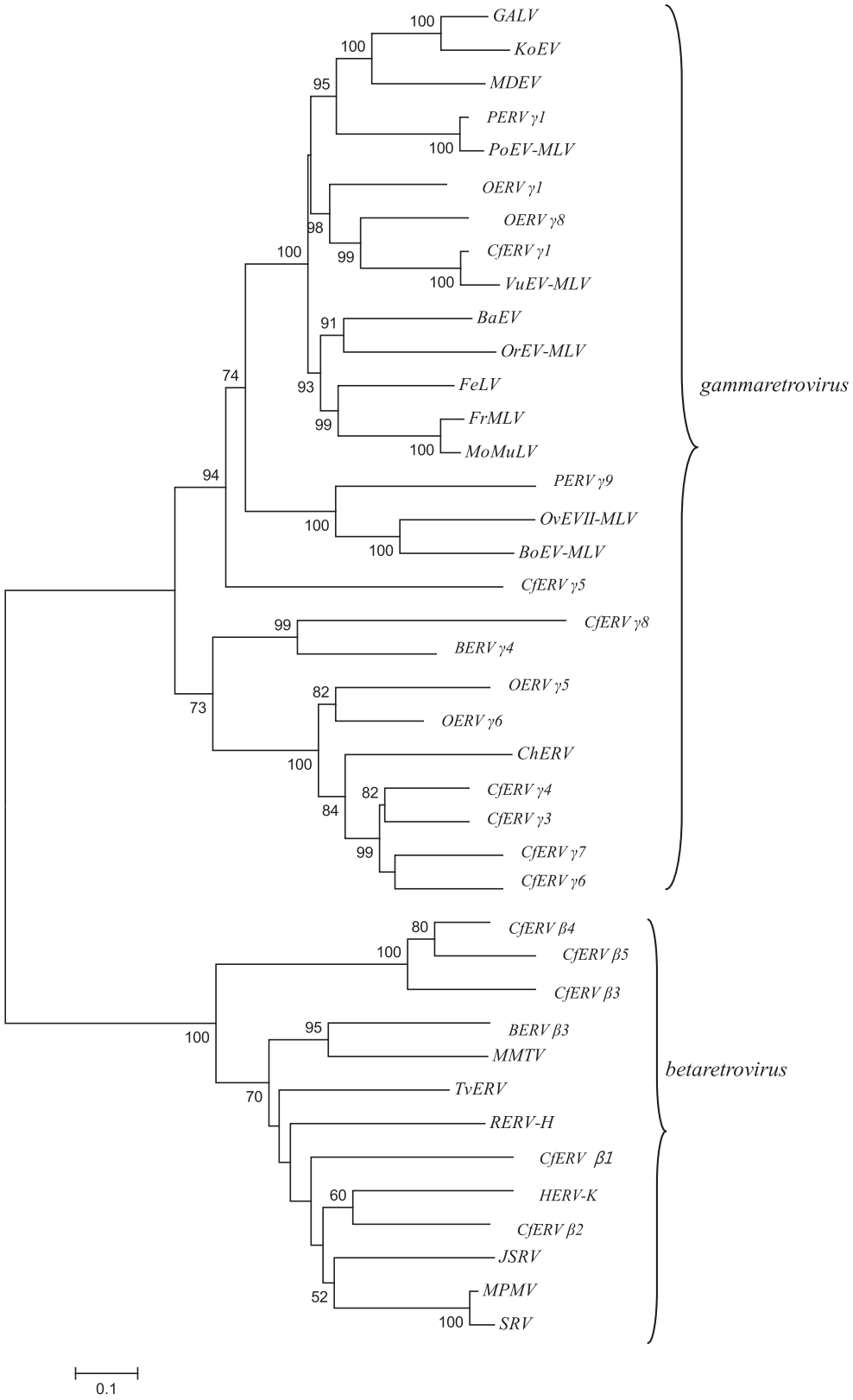


Fig. 6. Phylogenetic analysis using dog ERV elements of the PR–RT group with previously identified retroviruses. The *pro/pol* sequences with 814–943 bp were subjected to phylogenetic analysis using the neighbor-joining method. GenBank accession numbers for retroviral sequences are described in Materials and methods. Numbers at the branch nodes denote the bootstrap values (>50) from 1000 replicates. A representative sequence was used for each dog ERV family and the clone IDs are CfERVγ1 (*Der15-1*), CfERVγ3 (*Der6-1*), CfERVγ4 (*Der5-9*), CfERVγ5 (*Der1-12*), CfERVγ6 (*Der1-13*), CfERVγ7 (*Der1-7*), CfERVγ8 (*Der4-4*), CfERVβ1 (*Derb-32*), CfERVβ2 (*Derb-12a*), CfERVβ3 (*Derb-25*), CfERVβ4 (*Derb-27*), and CfERVβ5 (*Derb-x3*).

PRO (5′-GTK TTI KTI GAY ACI GGI KC-3′) and MLV-PRO (5′-YTI KTI GAY ACI GGI GCT SA-3′), and three reverse primers, CT (5′-AGI AGG TCR TCI ACR TAS TG-3′), JO (5′-ATI AGI AKR TCR TCI ACR TA-3′) and EM (5′-ATI AGI AKR TCR TCC ATR TA-3′). PCR products were gel purified using a QIAquick™ Gel Extraction Kit (QIAGEN, Germany) and ligated with pGEM-T Easy Vector (Promega, USA). The ligation products were electroporated into DH10B cells (Invitrogen, USA) using MicroPulser™ (Bio-Rad, USA). Transformed bacteria were plated onto agar containing 50 μg/ml ampicillin, 40 mg/ml X-gal, and 100 mM isopropyl β-D-1-thiogalactopyranoside (IPTG).

Sequencing

Plasmids were isolated using a GeneAll Exprep ™ Plasmid SV Kit (GeneAll Biotechnology, Korea). Sequencing reactions were performed using the ABI PRISM BigDye™ Terminator Cycle Sequencing Kit (Applied Biosystems, USA) with SP6 (5′-GAT TTA GGT GAC ACT ATA G-3′) and T7 (5′-TAA TAC GAC TCA CTA TAG GG-3′) universal sequencing primers in both forward and reverse directions. The full sequences of clone inserts corresponding to the *pro*/*pol* region were obtained by overlapping forward and reverse sequencing results. Ambiguous sequences or unique polymorphisms were confirmed by resequencing the same clone. Sequences were adjusted by eliminating vector and primer sequences.

In-silico identification of canine ERV

The dog genome assembly CanFam 2.0 (Lindblad-Toh et al., 2005) was used for the *in-silico* identification of ERVs in the dog genome. A previous study on the phylogenetic analysis of diverse RT-containing elements showed that sequences of retroviral origin can be differentiated from those of non-retroviral origin such as retrotransposons (Xiong and Eickbush, 1990). We combined the BLAST analysis results against the dog genome assembly from four different ERV sequence sources; 1) 81 *pro*/*pol* sequences (557–934 bp) obtained from the degenerative PCR, 2) three *Canis lupus familiaris* origin ERV sequences (~7000 bp) from the Repbase (http://www.girinst.org.repbase/), 3) *pro*/*pol* sequences of previously identified betaretroviruses (TvERV, SRV, JSRV, OERV β1, OERV β2, OERV β3, HERV-K, MPMV, PERV β5) and spuma-like ERV elements (MUERV-L and HERV-L) and 4) results of LTR_STRUC (McCarthy and McDonald, 2003) analysis using the canine genome. Redundant or severely mutated sequences were removed and individual ERV candidate sequences were reiteratively BLASTed against CanFam 2.0 until no new sequences were detected. Our matching criteria for the NCBI BLAST analysis was either an e-value$<10^{-25}$ or 60% nucleotide sequence identity with a minimum of 30% coverage.

Substructural characterization of canine ERVs

For the structural characterization of candidate ERV elements identified by the presence of *pro*/*pol* in the sequences, those corresponding to 7000–10,000 bp regions from both the 5′ and 3′ ends of the *pro*/*pol* sequences were aligned against one another using NCBI BLAST2SEQ to evaluate the presence of LTRs. The typical features of LTRs including the presence of di-nucleotide inverted repeat (TG-GA) (Baillie and Wilkins, 2001; Craven et al., 1995), polyadenylation signal, a TATA box, and a CCAAT box (Boonyaratanakornkit et al., 2004; Harada et al., 1987) were examined. We considered a minimum of 100 bp in length as possible LTR sequences. The candidate ERV elements were also examined by NCBI BLASTx to identify sequences corresponding to *gag*, *pro*, *pol*, and *env*. Amino acid motifs of the PR and RT region for each element were determined by NCBI BLASTx and multiple sequence alignment was carried out using the CLUSTALW2 EBI program (Chenna et al., 2003).

Nomenclature of canine ERVs

*Canisfamiliaris*endogenous retroviruses (CfERVs) were used to name ERVs in the dog genome. Because CfERV elements did not show significant sequence similarity to OERV, PERV, and BERV (BoERV) elements which shared interspecies sequence similarity among ERV families, we assigned the family names for the dog ERVs independent of previous studies of other species. The most abundant families from gamma- and betaretroviruses were referred to as CfERV γ1 and β1, respectively. Phylogenetic relationships to CfERV γ1 or β1 were used to name remaining ERV families.

Multiple sequence alignments, phylogenetic analysis, and estimation of proviral age

The alignments of the nucleotide or amino acid sequences were constructed using the CLUSTALW2 program (Chenna et al., 2003) and edited using CLC Main Workbench (http://clcbio.com). Phylogenetic analyses were performed using MEGA4.1 (Tamura et al., 2007). A neighbor-joining tree was built and bootstrap values were obtained from 1000 replicates. Sequence similarity within the family was estimated by pair-wise comparisons using BioEdit (Hall, 1999). Proviral age (T) was estimated according to the method used in previous studies (Johnson and Coffin, 1999; Tonjes and Niebert, 2003).

Nucleotide sequence accession numbers

A representative sequence each for CfERV γ1, γ2, γ3, and γ4 from degenerative PCRs were deposited to GenBank under accession numbers HM460338, HM460339, HM460340, and HM460341, respectively. Other GenBank accession numbers for reference sequences used in the analysis were Jaagsiekte Sheep Retrovirus (JSRV, M80216.1), mouse mammary tumor virus (MMTV, M15122.1), Simian type D virus 1 (SRV, U85505.1), Trichosurus vulpecula retrovirus (TvERV, AF224725.1), Mason-Pfizer monkey virus (MPMV, AAC82576.1), Ovisaries endogenous retrovirus (OERV) β1 (AY266332.1), OERV β2 (AY193894.1), OERV β3 (AY193895.1), rabbit endogenous retrovirus (RERV, AF480925), bovine endogenous retrovirus (BERV) β3 (DQ889607.1), porcine endogenous retrovirus (PERV) β5(AF511115.1) and human endogenous retrovirus (HERV-K, P10266.2) for betaretroviruses; murine leukemia related virus strain MeEVI, (MLV-MiERV1, X99926.1), PERVγ1 (AF511088.1), gibbon ape leukemia virus (GALV, AAA46810.1), feline leukemia virus (FELV, M18247.1), Friend murine leukemia virus (FrMLV, AAA46477.1), Moloney murine leukemia virus (MoMuLV, AF033811.1), Pan troglodytes polyprotein 1 (PtERV1, XM_001163307.1), Olive baboon endogenous retrovirus (BaEV, AJ507118.1), OERVγ8 (AY193908.1) and murine leukemia related virus strain VuEV (VuEV, X99935.1) for gammaretroviruses; and MuERV-L (Y12713.1) and HERV-L (X89211.1) for spuma-like ERVs.

Supplementary materials related to this article can be found online at doi:10.1016/j.virol.2011.10.010.

Acknowledgments

This work was supported by grants from the Biogreen 21 (No.20070401034029) and Next-Generation BioGreen 21 (No. PJ0080892011) Programs of the Rural Development Administration, Republic of Korea and also supported by the Konkuk University. We thank Dr. Goran Andersson and his colleagues at Swedish University of Agricultural Sciences for kindly providing us with the ERV dataset from their study.

References

Akiyoshi, D.E., Denaro, M., Zhu, H., Greenstein, J.L., Banerjee, P., Fishman, J.A., 1998. Identification of a full-length cDNA for an endogenous retrovirus of miniature swine. J. Virol. 72, 4503–4507.

Baillie, G.J., Wilkins, R.J., 2001. Endogenous type D retrovirus in a marsupial, the common brushtail possum Trichosurus vulpecula. J. Virol. 75, 2499–2507.
Benit, L., Dessen, P., Heidmann, T., 2001. Identification, phylogeny, and evolution of retroviral elements based on their envelope genes. J. Virol. 75, 11709–11719.
Blikstad, V., Benachenhou, F., Sperber, G., Blomberg, J., 2008. Evolution of human endogenous retroviral sequences: a conceptual account. Cell. Mol. Life Sci. 65, 3348–3365.
Blomberg, J., Benachenhou, F., Blikstad, V., Sperber, G., Mayer, J., 2009. Classification and nomenclature of endogenous retroviral sequences (ERVs): problems and recommendations. Gene 448, 115–123.
Boonyaratanakornkit, J., Chew, A., Ryu, D.D., Greenhalgh, D.G., Cho, K., 2004. Murine endogenous retroviruses and their transcriptional potentials. Mamm. Genome 15, 914–923.
Bowen, N.J., McDonald, J.F., 1999. Genomic analysis of *Caenorhabditis elegans* reveals ancient families of retroviral-like elements. Genome Res. 9, 924–935.
Bowen, N.J., McDonald, J.F., 2001. Drosophila euchromatic LTR retrotransposons are much younger than the host species in which they reside. Genome Res. 11, 1527–1540.
Britten, R.J., 1996. DNA sequence insertion and evolutionary variation in gene regulation. Proc. Natl. Acad. Sci. U. S. A. 93, 9374–9377.
Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., Thompson, J.D., 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31, 3497–3500.
Clutton-Brock, J., 1995. The Domestic Dog, its Evolution, Behavior and Interactions with People. Cambridge UniversityPress.
Coffin, J.M., Hughes, S.H., Varmus, H.E., 1997. The interactions of retroviruses and their hosts. In: Coffin, J.M., Hughes, S.H., Varmus, H.E. (Eds.), Retroviruses. CSHL Press, New York, pp. 335–341.
Craven, R.C., Leure-duPree, A.E., Weldon Jr., R.A., Wills, J.W., 1995. Genetic analysis of the major homology region of the Rous sarcoma virus Gag protein. J. Virol. 69, 4213–4227.
Dangel, A.W., Baker, B.J., Mendoza, A.R., Yu, C.Y., 1995. Complement component C4 gene intron 9 as a phylogenetic marker for primates: long terminal repeats of the endogenous retrovirus ERV-KC4. are a molecular clock of evolution. Immunogenetics 42, 41–52.
DeMartini, J.C., Carlson, J.O., Leroux, C., Spencer, T., Palmarini, M., 2003. Endogenous retroviruses related to jaagsiekte sheep retrovirus. Curr. Top. Microbiol. Immunol. 275, 117–137.
Dunwiddie, C.T., Resnick, R., Boyce-Jacino, M., Alegre, J.N., Faras, A.J., 1986. Molecular cloning and characterization of gag-, pol-, and env-related gene sequences in the ev- chicken. J. Virol. 59, 669–675.
Garcia-Etxebarria, K., Jugo, B.M., 2010. Genome-wide detection and characterization of endogenous retroviruses in Bos taurus. J. Virol. 84, 10852–10862.
Gifford, R., Tristem, M., 2003. The evolution, distribution and diversity of endogenous retroviruses. Virus Genes 26, 291–315.
Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.
Harada, F., Tsukada, N., Kato, N., 1987. Isolation of three kinds of human endogenous retrovirus-like sequences using tRNAPro. as a probe. Nucleic Acids Res. 15, 9153–9162.
Herniou, E., Martin, J., Miller, K., Cook, J., Wilkinson, M., Tristem, M., 1998. Retroviral diversity and distribution in vertebrates. J. Virol. 72, 5955–5966.
Huda, A., Polavarapu, N., Jordan, I.K., McDonald, J.F., 2008. Endogenous retroviruses of the chicken genome. Biol. Direct 3, 9.
Hughes, J.F., Coffin, J.M., 2005. Human endogenous retroviral elements as indicators of ectopic recombination events in the primate genome. Genetics 171, 1183–1194.
Jern, P., Sperber, G.O., Ahlsen, G., Blomberg, J., 2005. Sequence variability, gene structure, and expression of full-length human endogenous retrovirus H. J. Virol. 79, 6325–6337.
Johnson, W.E., Coffin, J.M., 1999. Constructing primate phylogenies from ancient retrovirus sequences. Proc. Natl. Acad. Sci. U. S. A. 96, 10254–10260.
Kimura, M., 1980. A simple method for estimating evolutionary ratesofbase substitutions through comparative studies of nucleotide-sequences. J. Mol. Evol. 16, 111–120.
Klymiuk, N., Muller, M., Brem, G., Aigner, B., 2002. Characterization of porcine endogenous retrovirus gamma pro-pol nucleotide sequences. J. Virol. 76, 11738–11743.
Klymiuk, N., Muller, M., Brem, G., Aigner, B., 2003. Characterization of endogenous retroviruses in sheep. J. Virol. 77, 11268–11273.
La Mantia, G., Maglione, D., Pengue, G., Dicristofano, A., Simeone, A., Lanfrancone, L., Lania, L., 1991. Identification and characterization of novel human endogenousretroviral sequences prefentially expressed in undifferentiated embryonal carcinomacells. Nucleic Acids Res. 19, 1513–1520.
Lebedev, Y.B., Belonovitch, O.S., Zybrova, N.V., Khil, P.P., Kurdyukov, S.G., Vinogradova, T.V., Hunsmann, G., Sverdlov, E.D., 2000. Differences in HERV-K LTR insertions in orthologous loci of humans and great apes. Gene 247, 265–277.
Lindblad-Toh, K., et al., 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438, 803–819.
Lower, R., Lower, J., Kurth, R., 1996. The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. Proc. Natl. Acad. Sci. U. S. A. 93, 5177–5184.
Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., van de Lagemaat, L.N., Mager, D.L., 2006. Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS Genet. 2, e2.
Martínez Barrio, Á., Ekerljung, M., Jern, P., Benachenhou, F., Sperber, G.O., et al., 2011. The first sequenced carnivore genome shows complex host–endogenous retrovirus relationships. PLoS One 6 (5), e19832.
McCarthy, E.M., McDonald, J.F., 2003. LTR_STRUC: a novel search and identification program for LTR retrotransposons. Bioinformatics 19, 362–367.
McCarthy, E.M., McDonald, J.F., 2004. Long terminal repeat retrotransposons of Mus musculus. Genome Biol. 5, R14.
McCarthy, E.M., Liu, J., Gao, L., McDonald, J.F., 2002. Long terminal repeat retrotransposons of *Oryza sativa*. Genome Biol. 3 research0053.1-0053.11.
McDonald, J.F., 1993. Evolution and consequences of transposable elements. Curr. Opin. Genet. Dev. 3, 855–864.
Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16, 1215.
Palmarini, M., Murgia, C., Fan, H., 2002. Spliced and prematurely polyadenylated Jaagsiekte sheep retrovirus-specific RNAs from infected or transfected cells. Virology 294, 180–188.
Patience, C., Switzer, W.M., Takeuchi, Y., Griffiths, D.J., Goward, M.E., Heneine, W., Stoye, J.P., Weiss, R.A., 2001. Multiple groups of novel retroviral genomes in pigs and related species. J. Virol. 75, 2771–2775.
Patterson, D.F., Haskins, M.E., Jezyk, P.F., 1982. Models of human genetic disease in domestic animals. Adv. Hum. Genet. 12, 263–339.
Patterson, D.F., et al., 1988. Research on genetic diseases: reciprocal benefits to animals and man. J. Am. Vet. Med. Assoc. 193, 1131–1144.
Polavarapu, N., Bowen, N.J., McDonald, J.F., 2006. Identification, characterization and comparative genomics of chimpanzee endogenous retroviruses. Genome Biol. 7, R51.
Ryan, F.P., 2004. Human endogenous retroviruses in health and disease: a symbiotic perspective. J. R. Soc. Med. 97, 560–565.
Schwartz, M., 1997. History of Dogs in the Early Americas. Yale University Press.
Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis MEGA. software version 4.0. Mol. Biol. Evol. 24, 1596–1599.
Tonjes, R.R., Niebert, M., 2003. Relative age of proviral porcine endogenous retrovirus sequences in Sus scrofa based on the molecular clock hypothesis. J. Virol. 77, 12363–12368.
Tristem, M., 1996. Amplification of divergent retroelements by PCR. Biotechniques 20, 608–612.
Tristem, M., Kabat, P., Lieberman, L., Linde, S., Karpas, A., Hill, F., 1996. Characterization of a novel murine leukemia virus-related subgroup within mammals. J. Virol. 70, 8241–8246.
Vogt, V.M., 1997. Retroviral virions and genomes. In: Coffin, J.M., Hughes, S.H., Varmus, H.E. (Eds.), Retroviruses. CSHL Press, New York, pp. 335–341.
Wayne, R.K., Ostrander, E.A., 1999. Origin, geneticdiversity, and genome structure of the domestic dog. Bioessays 21, 247–257.
Weiss, R.A., 2006. The discovery of endogenous retroviruses. Retrovirology 3, 67.
Xiao, R., Park, K., Oh, Y., Kim, J., Park, C., 2008a. Structural characterization of the genome of BERV gamma4, the most abundant endogenous retrovirus family in cattle. Mol. Cells 26, 404–408.
Xiao, R., Park, K., Lee, H., Kim, J., Park, C., 2008b. Identification and classification of endogenous retroviruses in cattle. J. Virol. 82, 582–587.
Xiong, Y., Eickbush, T.H., 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9, 3353–3362.